International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I8P111

Volume 73 Issue 8, 129-136, August 2025

Original Article

Next-Gen Air Quality Index Forecasting with Hybrid Machine Learning Models and Cloud Synergy

Diana George[1], R. Navya[2], Vinitha V[3]

[1]*Department of Computer Science and Engineering, Dayananda Sagar University, Bangalore, Karnataka, India.*
[2]*Department of Electronics and Communication Engineering, Dayananda Sagar University, Bangalore, Karnataka, India.*
[3]*Department of Cyber Security, Dayananda Sagar University, Bangalore, Karnataka, India.*

[1]*Corresponding Author : dianageorge-cse@dsu.edu.in*

Received: 01 March 2025 Revised: 17 July 2025 Accepted: 31 July 2025 Published: 30 August 2025

Abstract - *Globally, nowadays, air pollution remains a major menace in terms of both environmental and public health; as such, accurate monitoring and forecasting the quality of air are essential for mitigating its deleterious impact. The Air Quality Index (AQI) is used to detect the quality of air and its hazardous effects on human health. This paper tries to formulate a forecasting mechanism for AQI by measuring the rate of the major issues causing air pollutants such as PM2.5, PM10, O3, CO, SO2, NO2, Pb(lead), and NH3. Hence, this paper formulates a model that combines both Convolutional Neural Networks (CNNs) along with Transformers and an enhanced Attention Mechanism to improve the prediction accuracy. CNNs are intended for effective feature extraction and capturing spatial patterns in air quality data, while the transformer model captures the sequential dependencies, allowing for accurate predictions over time. This proposed hybrid model addresses the limitations of age-old time-series models like ARIMA and LSTM, which often struggle to analyze the complex spatial-temporal air quality relationship. The proposed model was trained using the same historical air quality data provided by the Government of India (GoI) for training and validation, with real-time deployment with live sensor data. Also, the use of cloud computing ensures efficient handling of live data streams, enabling real-time data processing, prediction, and updates. This allows for quick, scalable and reliable predictions on large, diverse datasets, timely public health alerts, and supports proactive environmental management.*

Keywords - *Air Quality Forecasting, Attention mechanism, Transformer, CNN, Cloud computing.*

1. Introduction

It is an inevitable fact that air pollution is one of the major environmental hazards affecting global health. With increasing urbanization and industrialization, accurate forecasting of air quality has become imperative for governments and other organizations to manage and track the pollution levels effectively. The Air Quality Index (AQI) is used to categorize air pollution levels of different pollutants based on several attributes and provide recommendations to the public. This paper proposes the integration of Attention Mechanisms, Transformers, and Convolutional Neural Networks (CNNs) for more accurate AQI predictions. Traditional forecasting models such as ARIMA, Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks have limitations, including difficulty handling long-range dependencies and the need for feature engineering. Transformers, with their self-attention mechanisms, and CNNs, with their ability to capture spatial patterns, provide an excellent solution for these issues. By combining these two methods, this work aims to capture both spatial and temporal dependencies in air quality data, significantly improving forecasting accuracy.

However, models such as ARIMA and LSTM have some limitations when dealing with complex air quality data. ARIMA often shows difficulty in capturing non-linear relationships. Despite the capability of LSTM in learning time-based patterns, it sometimes struggles with long-term dependencies as well as spatial intricacies. On top of this, there was a notable lack of integrated approaches that leverage extended deep learning techniques with real-time data processing with the support of cloud infrastructure. So, this study aims to fill this gap by formulating a hybrid approach that utilizes the spatial learning ability of Convolutional Neural Networks, the Transformer's temporal modelling strength, and the option of cloud computing for scalability.

Moreover, the integration of the proposed system with AWS cloud services ensures scalable computation and real-time data ingestion and prompts AQI forecasting across diverse geographic regions. Hence, it is evident that this newly proposed end-to-end architecture is the first to holistically combine deep learning algorithms, attention mechanisms, and cloud computing for dynamic AQI monitoring by surpassing the limitations of existing standalone models.

2. Literature Survey

Research related to the quality of air and its forecasting has evolved through several modelling approaches, stretching from traditional statistical models to the advanced state-of-the-art deep learning frameworks. This section presents the major contributions to ARIMA-based, LSTM-based, Transformer-based, hybrid, and cloud-enabled real-time models.

2.1. ARIMA-Based Models

The ARIMA is one of the earliest statistical tools employed for time-series forecasting in environmental studies due to its efficiency in modelling short-term linear trends in pollutant concentration. Nonetheless, its ability to capture non-linear and rapidly changing patterns remains a question. To exemplify, studies like that of X. Yuan et al. [1] used ARIMA with decomposition and other filtering methods to enhance prediction accuracy. Despite improvements, ARIMA alone is not capable of dealing with multi-dimensional and fluctuating air quality data.

2.2. LSTM-Based Models

To overcome ARIMA's limitations in handling temporal non-linearity, Long Short-Term Memory (LSTM) networks have been introduced extensively. These models are well-suited for capturing sequential patterns and temporal dependencies in pollution data. Z. Zhou [2] demonstrated how a Bayesian variant of LSTM could predict PM2.5 levels with high accuracy while offering interpretability. Similarly, Preethi et al. [3] incorporated Bi-LSTM into an IoT-enabled framework, allowing for real-time predictions. Even though it is powerful, LSTM models tend to require substantial training data and sometimes fail to capture long-range dependencies in highly unstable environments.

2.3. Transformer-Based Models

Transformer models, initially formulated for language processing, have shown significant relevance in time-series forecasting due to their self-attention mechanisms. Unlike LSTM, Transformers are better at obtaining long-term dependencies without relying on recurrence. Q. Guo et al. [4] proposed a Transformer-based AQI predictor, Air Former, that outperformed several traditional models across extensive datasets. In addition, Wu J et al. [5] presented a framework incorporating interpretability and uncertainty estimation, which is crucial for policy-level decisions. However, these models tend to focus more on temporal aspects and often miss spatial context, which is also vital in air quality analysis.

2.4. Hybrid Models

In present times, many hybrid models achieve better results than relying on a single technique due to the contribution of the advantages of every single unit. It ultimately integrates statistical and neural components, which have shown superior accuracy in forecasting. Yanrong Ma et al. [6] used a combination of SARIMA and LSTM, optimized via the Sparrow Search Algorithm, to capture both trend and noise in AQI patterns. Dong Z et al. [7] took a broader approach by combining decomposition, optimization, and machine learning elements into a unified hybrid system. While such models improve predictive strength, they can be computationally intensive and sometimes lack portability for real-time applications.

2.5. Real-Time and Cloud-Based Forecasting

Due to the increasing demand for responsive air quality systems, integrating machine learning with cloud and IoT technologies has become essential. Preethi et al. [3] demonstrated how a Bi-LSTM-based model embedded in an IoT system could offer prompt AQI forecasts. Further innovations like those by S. Veera Manikandan et al. [8] have explored IoT-based frameworks leveraging fog computing for real-time pollutant analysis at the edge. In another effort, N. Nilesh et al. [9] curated a large-scale real-time AQI dataset for machine learning model training. These real-time systems are promising, yet they face challenges in terms of data consistency, sensor calibration, and generalizability across different regions.

3. Methodology

The methodology outlines the step-by-step procedures for the development of a hybrid model for predicting air quality index (AQI) using CNN with Transformer's Self-Attention Mechanisms. The key steps in the process are explained in detail in the following sessions.

3.1. Data Set Collection

The first step is to gather the air quality data from real-time sensors and historical datasets. These datasets typically contain pollutant rates of PM2.5, PM10, SO2, O3, CO, NO2, Pb, and NH3, along with the parameters values that affect the pollutant values. The Parameters that has taken into account for the implementation of this model are Temperature, Minimum Temperature (Tm), Moisture Content (PP), Visibility Value (VV), Humidity(H), Wind Rate(V), Maximum Wind Rate (VM) timestamps, Maximum Temperature (TM), and other metadata like geographic location from government sensors, environmental monitoring systems.

Table 1. Attribute list considered for this model training

Attribute	Details
Source of the Data	Government of India (GoI) Air Quality Data
Features	PM2.5, PM10, O3, CO, NO2, SO2, Pb, NH3
Data Size	X observations (e.g., number of records)
Sensor Locations	Various locations in India

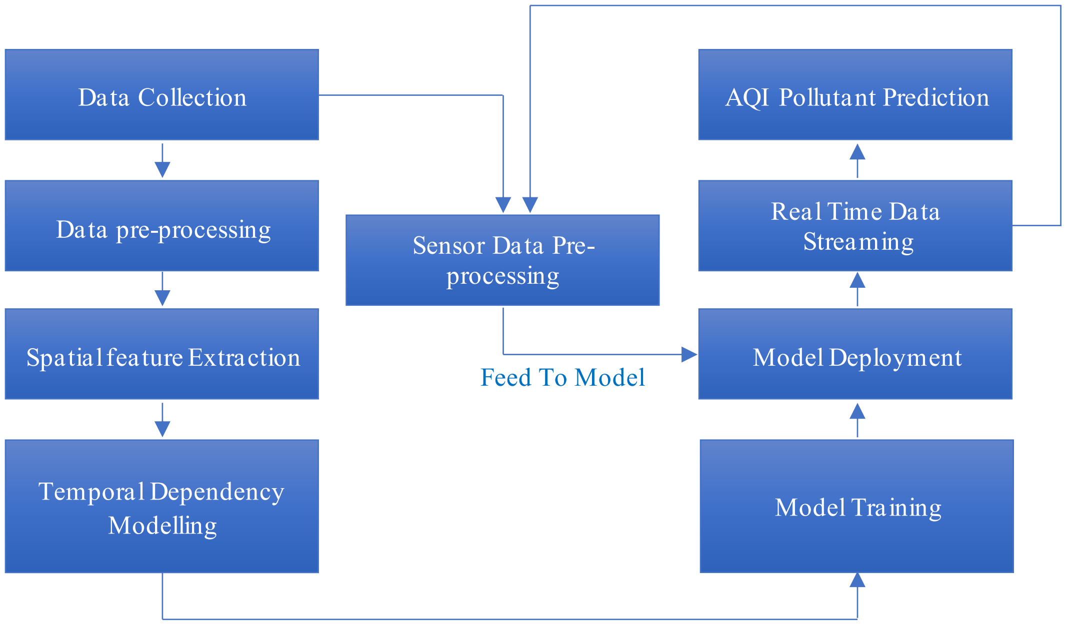

Fig. 1 Workflow of the proposed model

3.2. Data Pre-Processing

This is an essential step before applying the model to ensure the data is clean, normalized, and well-represented.

3.2.1. Normalization

Normalization ensures that each pollutant feature is scaled between the range of 0 and 1. This can be ensured through Min-Max normalization. Given a raw pollutant value x_{raw}, the normalized value x_{norm}, is calculated as:

$$x_{norm} = \frac{x_{raw} - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Where,

- x_{raw} is the raw pollutant value.
- x_{min} is the minimum and x_{max} is the maximum value of the feature of the dataset.
- x_{norm} is the normalized value of the pollutant.

3.2.2. Handle Data Missing

To handle the missing data in the dataset, the model utilizes the advantages of the moving average method. For a pollutant x_i at time step i, the imputed value $X(imputed)_i$ is computed as the average of the previous k and next k neighbouring values:

$$X(imputed)_i = \frac{1}{2k+1} \sum_{j=i-k}^{i+k} x_j \tag{2}$$

Where,

- x_j is termed as the observed pollutant values at neighbouring time steps,
- K is known as the window size, typically set to 3 or 5 based on the dataset.

3.2.3. Feature Engineering (For Temporal Features)

In this step, temporal features like time of day, day of the week, and the month of the year need to be captured to find the periodicity of air quality data.

- Time of a Day (Hour): Extracting the hour from the timestamp T using

  $$\mathrm{Hour}(\mathrm{T}) = \mathrm{T} \bmod 24 \tag{3}$$

  Where, Hour(T) is the hour of the day, ranging from 0 to 23.

- Day of the Week: Extracting the day of the week from the timestamp using

  $$\mathrm{DayOfWeek}(\mathrm{T}) = \left\lfloor \frac{\mathrm{T}}{24 \times 7} \right\rfloor \tag{4}$$

  Where, DayOfWeek(T) is an integer from 0 (Sunday) to 6 (Saturday)

- Month of a Year: Extracting the month of the year using

  $$\mathrm{Month}(T) = \left\lfloor \frac{\mathrm{T}}{24 \times 30} \right\rfloor + 1 \tag{5}$$

  Where, Month(T) is an integer from 1 (January) to 12 (December).

These features help the model capture the seasonal and periodic patterns of the air quality data.

3.3. Spatial Feature Extraction By CNN

Spatial pattern features from the air quality data extracted by the CNN are happened in this phase. Hence, given an input feature matrix $x_{input,}$ representing the air quality data, where each feature corresponds to a pollutant at a specific location, the CNN model applies convolutional filters W to capture spatial features.

$$Y_{Feature} = Conv2D\,(x_{input,} W, b) \qquad (6)$$

Where,

- $x_{input,}$ is the raw input feature matrix,
- W is the convolutional kernel (filter),
- b is the bias term,
- $Y_{Feature}$ is the output feature map from the convolution.

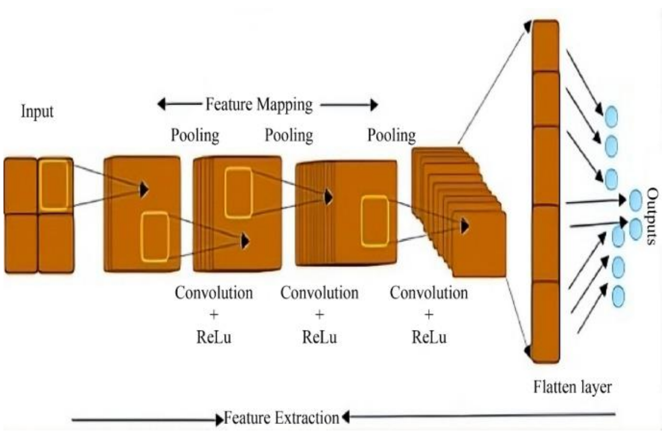

Fig 2. Architecture of CNN-feature extraction

After convolution, a non-linear activation function (ReLU) is applied:

$$Y_{Activation} = \mathrm{ReLU}(Y_{Feature}) \qquad (7)$$

The feature map generated by the CNN captures spatial patterns in the air quality data, which are then passed to the Transformer for temporal modelling.

3.4. Transformer for Temporal Sequence Modelling

Once spatial features are extracted, the role of the transformer will come into play for capturing temporal dependencies from the pollutant value. The input to the Transformer is the feature map. $Y_{Activation}$ from the CNN. The input to the Transformer is transformed using an embedding layer:

$$X(transformer)_{input} = Y_{Activation} \cdot W_{Embedding} \qquad (8)$$

Where,

- $W_{Embedding}$ is the embedding matrix that maps the spatial features into a higher-dimensional space.

The Transformer model employs self-attention to model long-range dependencies between different time steps. The self-attention mechanism calculates attention scores as:

$$Attention\,(Q, K, V) = \frac{QK^T}{\sqrt{d_k}} \qquad (9)$$

Where,

- Q is the query, K is termed the key and V is mentioned as the value matrices.
- d_k is the key vector dimension.

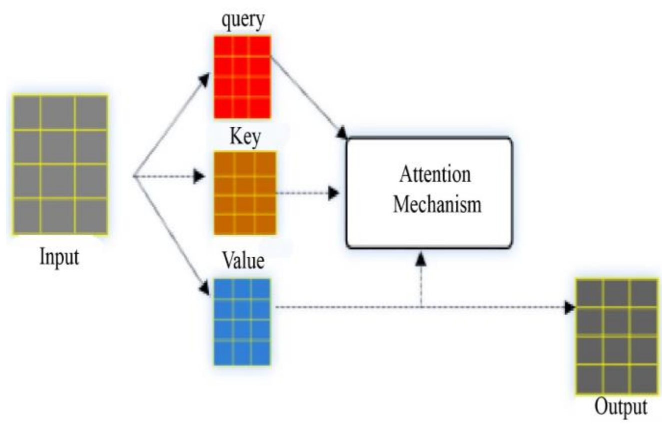

Fig 3. Transformer's attention mechanism

The output of it is a weighted sum of the value vectors:

$$Y_{Activation} = \sum_{t=1}^{T} (\alpha_t \,.\, Y_t) \qquad (10)$$

Where,

- α_t is the attention weight at time t
- Y_t is the predicted output at time t

This attention mechanism allows the model to focus on relevant time steps and improve prediction accuracy.

3.5. Model Training

The model is trained and evaluated by the Mean Squared Error (MSE) loss function that minimizes the difference (if any) between predicted and actual AQI by considering the following hyperparameter values.

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^{N} \left(Y_{pred,i} - Y_{true,i}\right)^2 \qquad (11)$$

Here,

- $Y_{pred,i}$ is termed as the predicted value at the time step i,
- $Y_{true,i}$ is known for the actual value at the time step i,
- N is the total number of time steps.

Table 2 will provide a clear insight into various hyperparameter values for the best accuracy of the proposed model.

Table 2. Hyperparameter values need to be set for the model

Hyperparameter	Magnitude
Optimizer	Adam gradient descent optimizer
Batch Size	32
Learning Rate	0.001
Epochs	50
Dropout Rate	0.3

3.6. Performance Evaluation

The adaptation of the model is assessed using several key metrics, namely:

- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_{pred,i} - Y_{true,i}\right)^2} \quad (12)$$

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|Y_{pred,i} - Y_{true,i}\right| \quad (13)$$

These metrics provide a clear picture of how accurately the model predicts AQI values.

3.7. Model Deployment in the Cloud

The model that has been trained using the static historical data is then deployed and hosted in the cloud platform, i.e., on AWS, and now the web service can accept the sensor real-time data and is able to make accurate predictions based on it.

3.8. Real-Time Data Processing

Once the model is deployed to the cloud, it will stream live sensor data. For easy analysis, the whole process from here onwards is broken down into numerous Stages.

3.8.1. Data Ingestion using AWS Kinesis

The live air related data from various air quality sensors like pollutant values (e.g., PM2.5, PM10, CO, NO2, SO2, etc.), sensor metadata (such as Sensor ID, Location, Timestamp), and environmental parameters (Temperature(T), Maximum Temperature (TM), Minimum Temperature (Tm), Humidity(H), Moisture Content (PP), Visibility Value (VV), Wind Rate(V), Maximum Wind Rate (VM)), used to send to the cloud through AWS Kinesis pipelining. This AWS Kinesis plays the role of a real-time data ingestion platform, providing high-throughput and low-latency data transfer to the cloud environment.

3.8.2. Data Pre-processing, Feature Engineering and Normalization

After obtaining the data, it needs to undergo a cleaning and pre-processing stage for the exact prediction of air quality by the deployed model. It involves handling missing or corrupted data, outlier detection, and ensuring the data format is consistent. Then, Feature Engineering and Data normalization should be done respectively on the processed data by automating the Cloud services such as AWS Lambda, because it enables serverless computing, allowing for scalability and low operational overhead. Feature engineering targets to enhance the input data for the prediction model by deriving additional features from the raw data, such as temporal features as well as aggregating pollutant data from different sensors based on geographical proximity to create meaningful inputs for the model, However, Normalization is a pivot step as it standardizes the pollutants' values and ensure they are on a comparable scale before feeding them into the model for prediction.

3.9. AQI Prediction

After processing, the data is pipelined to the Transformer model and attention mechanism, and the model generates the final AQI prediction. This is typically done by aggregating the contributions from different pollutants, as AQI is a composite index. As per the Government of India, AQI is categorized into various ranges based on the value of several listed pollutants. Such values are formulated and shown in the table below.

As such, AQI is calculated by combining pollutant concentrations and applying a set of weights:

$$AQI = \sum_{i=1}^{n}\left(w_i . Pollutant_i\right) \quad (14)$$

Where,

- w_i is the weight for pollutant i,
- $Pollutant_i$ is the concentration of pollutant i,
- n is the number of pollutants considered.

4. Results Analyses and Discussion

4.1. Results and Comparative Analysis

To validate the performance, it is required to compare the obtained result against several other baseline models, like ARIMA, LSTM, RF, and SVM, for better clarity and confirmation of the superior performance. First, the model was evaluated based on several key error metrics as mentioned in Table 3. These metrics ultimately provide a comprehensive understanding of prediction accuracy, model generalization, and consistency.

Table 3. Performance evaluation and comparison of various models based on AQI prediction

Model	RMSE (µg/m³)	MAE (µg/m³)	R² Score (%)
ARIMA	24.51	18.84	72.23
LSTM	17.78	12.42	85.11
Random Forest	15.65	11.07	87.04
Support Vector Machine (SVM)	19.29	13.83	77.56
Proposed Hybrid CNN + Transformer	8.32	6.88	95.37

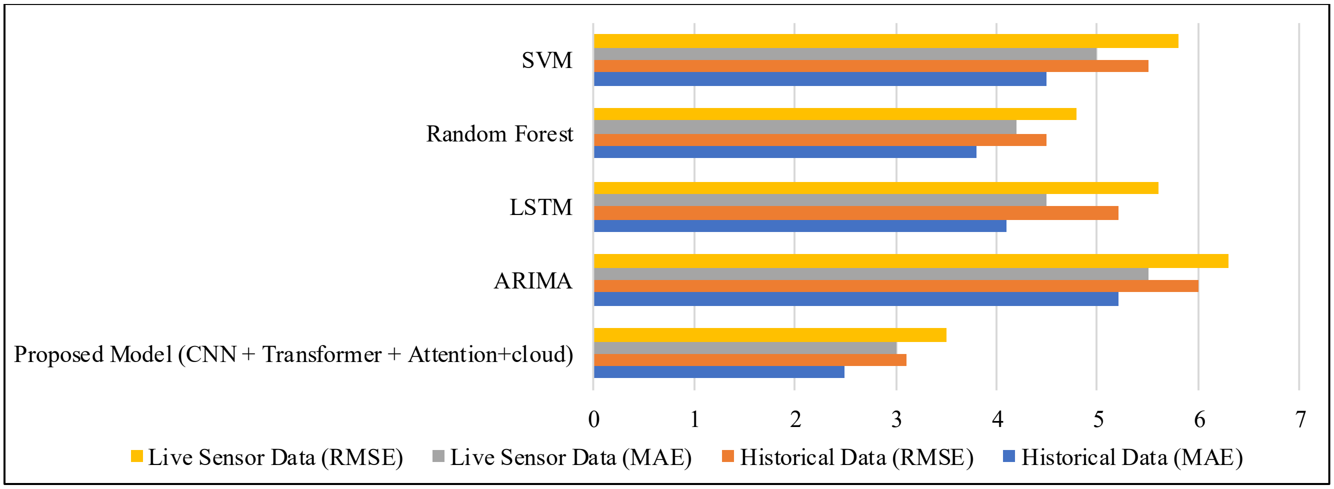

Fig. 4 RMSE Trends over Forecast Days

The results clearly demonstrate that the proposed hybrid model undoubtedly reduces prediction errors and achieves the highest R^2 score among the contemporaries, indicating a strong correlation between actual and predicted AQI values.

4.2. Temporal Error Trend Analysis

To further evaluate the model's robustness, the RMSE and MAE values were calculated over a weekly prediction interval. The graph below clearly illustrates the fluctuations in RMSE across a 14-day forecast window using different models.

As compared to the baseline models, the proposed hybrid model upheld a steadily low RMSE over time, meaning better stability as well as generalization capability. In contrast, traditional models like ARIMA showed higher variance, especially when pollutant levels were highly dynamic.

4.3. Performance Analysis Based on Region

The model was tested with the datasets collected from urban, suburban, and rural environments. The results had shown slightly better performance.

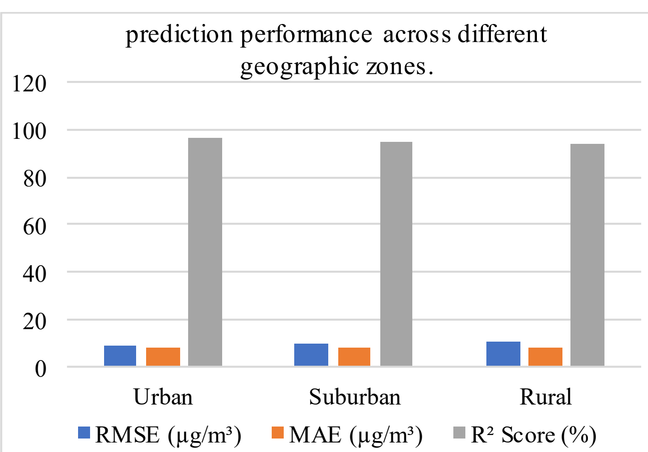

Fig. 5 AQI prediction performance across different geographic zones

4.4. Comparison Discussion

It is evidently proven that the proposed model outperforms the existing conventional methods due to the complementary strengths of individual components. Models like ARIMA are effective only in capturing linear patterns but fail in dealing with intrinsic non-linear relationships in air quality data. While LSTM networks are capable of modelling temporal dependencies, they show low performance with long-range sequences and lack spatial awareness.

However, the Convolutional Neural Network (CNN) component in the proposed model captures spatial variations of pollutant concentrations across different regions. This allows the system to learn localized air quality patterns more effectively. In addition to that, the Transformer network, with an enhanced self-attention mechanism, drives long-range temporal dependencies far better than LSTM, enabling the model to consider pollutant behaviour across extended time periods.

Also, a key game-changer of the proposed model is the deployment to the cloud platform for real-time prediction by allowing continuous ingestion of live sensor data, scalable processing, and immediate AQI prediction and alert generation. It significantly improves the model's practical usability for environmental monitoring and public health response awareness.

4.5. Limitations and Future Work

Even though the proposed model depicts superior performance in forecasting AQI, certain limitations still exist, like its accuracy heavily relies on the quality and availability of sensor data, which may or may not vary across regions. Additionally, deploying the system on the cloud infrastructure sometimes introduces latency issues, leading to higher costs in remote or resource-constrained environments. Most prominently, the current model does not account for some other external factors, such as traffic density or industrial emissions. In future work, as an extension, it is planned to experiment on satellite-based environmental data and expand its adaptability to accommodate different national AQI standards. Moreover, the development of mobile-based AQI alert systems is planned to improve public accessibility and real-time responsiveness.

5. Conclusion

Air pollution remains a challenge that needs to be tackled on an urgent basis because it creates significant risks to public health, besides posing an environmental threat. Therefore, sound forecasting of the air quality is requisite for mitigating these challenges. Hence, this paper attempts to introduce an advanced state-of-the-art forecasting model that integrates Convolutional Neural Networks (CNNs), Transformer networks enhanced by a self-attention mechanism, and cloud computing infrastructure for better accuracy while dealing with sophisticated spatial and temporal patterns of the air.

The CNN component of the model excels at capturing spatial dependencies by processing and extracting features from air quality data, and these features are vital for estimating how air quality varies across diverse environments for the accurate prediction of the Air quality of a given locality.

On the temporal side, the Transformer network with a self-attention mechanism plays a crucial role in the effective capture of long-range dependencies within the data, as this result will help in understanding past pollutant behaviours trend and how it influences future forecasts.

The real-time prediction process is further made reliable and accurate by the deployment of the model on cloud infrastructure. Cloud computing provides the room for handling large-scale datasets, especially live continuous streams of sensor data; hence, it ensures that the model can operate in real-time and is capable of providing up-to-date predictions. The cloud deployment also ensures scalability, as the model can easily handle data from multiple air sensors across different regions.

To conclude, this model combines the strengths of CNNs for spatial pattern extraction, Transformers for modelling complex temporal dependencies, and cloud infrastructure for real-time, scalable processing. Through addressing the spatial-temporal challenges in air quality forecasting, the model provides a novel, robust solution for predicting AQI and enabling timely health alerts. This fine, innovative approach can significantly improve air quality monitoring systems.

References

[1] Xiqian Yuan et al., "Air Quality Secondary Forecast Model based on ARIMA Time Series Model," *2023 3rd Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)*, Shenyang, China, pp. 310-313, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Ziqi Zhou, "Air Quality Prediction Based on Improved LSTM Model," *2023 4th International Conference on Computer Engineering and Application (ICCEA)*, Hangzhou, China, pp. 392-395, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] P. Preethi et al., "A Real-time Environmental Air Pollution Predictor Model Using Dense Deep Learning Approach in IoT Infrastructure," *Global NEST Journal*, vol. 26, no. 3, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Qiwei Guo et al., "Air Quality Time Series Analysis and Prediction Based on SMA-Transformer-QR," *2025 IEEE International Conference on Electronics, Energy Systems and Power Engineering (EESPE)*, Shenyang, China, pp. 1572-1576, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] Junhao Wu et al., "A Novel Framework for High Resolution Air Quality Index Prediction with Interpretable Artificial Intelligence and Uncertainties Estimation," *Journal of Environmental Management*, vol. 357, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yanrong Ma, Jun Ma, and Yifan Wang et al., "Hybrid Prediction Model of Air Pollutant Concentration for $PM_{2.5}$ and PM_{10}," *Atmosphere*, vol. 14, no. 7, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Zefan Dong, and Yonghui Zhou, "A Novel Hybrid Model for Financial Forecasting Based on CEEMDAN-SE and ARIMA-CNN-LSTM," *Mathematics*, vol. 12, no. 16, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] S. Veera Manikandan et al., "Internet of Things Enabled ML for Air Quality Assessment: Systematic Review," *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, pp. 1509-1514, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Nitin Nilesh et al., "IoT and ML-based AQI Estimation using Real-time Traffic Data," *2022 IEEE 8th World Forum on Internet of Things (WF-IoT)*, Yokohama, Japan, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] S. Mathupriya et al., "Localized Detection, Control and Combating of Greenhouse Gas Emissions using IoT and Data Science," *2024 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, Chennai, India, pp. 1-5, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Fabian Surya Pramudya, Felix Indra Kurniadi, and Aldilla Noor Rakhiemah, "Breathing in Jakarta: Uncovering the Air Quality Index using Data Visualization," *2023 7th International Conference on New Media Studies (CONMEDIA)*, Bali, Indonesia, pp. 162-166, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Snehal Lawande et al., "Comparative Analysis of Machine Learning Models for Prediction of Air Quality Index," *2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA)*, Pune, India, pp. 1-5, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Vanshay Gupta, Samit Kapadia, and Chetashri Bhadane, "Time Series Analysis and Forecasting of Air Quality in India," *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Erode, India, pp. 1-5, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[14] Xiaonan Sun, and Yundong Gu, "Evaluation of Ambient Air Quality Grade Based on Cloud Model and Combination Weighting Method," *2024 7th International Conference on Computer Information Science and Application Technology (CISAT)*, Hangzhou, China, pp. 1007-1010, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Abhishek Walia et al., "Prediction of Air Quality Index Using Random Forest and Prophet Tool," *2024 19th Annual System of Systems Engineering Conference (SoSE)*, Tacoma, WA, USA, pp. 275-280, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Swati Jha et al., "IoT Based Incident Control System using Air Quality Index Monitoring System," *2022 International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS)*, Nagpur, India, pp. 108-112, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Sandeep Kumar Sunori, Pushpa Bhakuni Negi, and Pradeep Juneja, "Estimation of Air Quality Index using AI and ML Techniques," *2023 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, Theni, India, pp. 1078-1082, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Nanda Aulia Sofiah et al., "A Comparative Assessment SARIMA and LSTM Models for the Gurugram Air Quality Index's Knowledge Discovery," *2024 International Conference on Electrical Engineering and Computer Science (ICECOS)*, Palembang, Indonesia, pp. 26-31, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Swati Varshney, Jitendra Nath Shrivastava, and Neha Gupta, "Machine Learning Algorithms for Forecasting Air Quality Index: A Predictive Analysis in the Taj Trapezium Zone (TTZ) of Agra," *2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India, vol. 3, pp. 1-6, 2025. [CrossRef] [Google Scholar] [Publisher Link]